

WITOLD KIERAŚ

SCHWYZERTÜÜTSCH, BAMBARA I JĘZYKI BEZKONTEKSTOWE

I. WPROWADZENIE

Noam Chomsky w swojej często cytowanej, ale rzadko czytanej pracy (Chomsky 1956) zaproponował hierarchię gramatyk formalnych rozumianych jako zbiory reguł przepisywania ciągów znaków terminalnych i nieterminalnych na inne ciągi znaków terminalnych i nieterminalnych. Dziś powszechnie znana i stosowana jest pod nazwą hierarchii Chomsky'ego. Co prawda autor, jak sam twierdzi, opisuje w ten sposób jedynie języki formalne, jednak na marginesie rozważań sformułował problem, który dotyczy języków naturalnych: do jakiej klasy gramatyk formalnych należeć będą gramatyki adekwatnie opisujące języki naturalne? Sam nie udzielił na to pytanie odpowiedzi, ale wzbudził nim długą dyskusję filozoficzną i lingwistyczną na temat złożoności konstrukcji w różnych językach naturalnych.

Już na samym początku Chomsky odrzucił dwie klasy swojej hierarchii, jako nieadekwatne do opisu gramatyk języków naturalnych: gramatyki typu 0 (gramatyki nieograniczone, generujące języki rekurencyjnie przeliczalne) oraz gramatyki typu 3 (gramatyki regularne). Pierwszą z klas uznał za zbyt szeroką – informacja, że języki naturalne można opisywać za pomocą gramatyk nieograniczonych, które potrafią generować dowolny język formalny, przy jednoczesnym założeniu, że adekwatnego opisu języków naturalnych można dokonać metodami czysto formalnymi (a takie założenie niewątpliwie Chomsky czyni), jest właściwie bezużyteczna. Drugą z klas – gramatyki regularne – Chomsky odrzuca jako mechanizm zbyt słaby, na co wskazywał również w napisanej wcześniej, ale opublikowanej dopiero w 1957 roku książce: *Syntactic Structures* (Chomsky 1957). Zatem postawiony przez Chomsky'ego problem dotyczy tylko dwóch klas: gramatyk kontekstowych i gramatyk bezkontekstowych.

2. MOTYWACJE

Istnieje kilka powodów, dla których pytanie, czy gramatyki bezkontekstowe są wystarczającym mechanizmem do opisania wszystkich języków naturalnych, jest interesujące.

Pierwszy powód jest czysto praktyczny: w wypadku ogólnym problem rozpoznawania (i parsowania) języków bezkontekstowych jest wielomianowy, czyli obliczalny w sensie praktycznym. W wypadku języków kontekstowych jest to jednak problem obliczeniowo trudny. Zatem w obliczu wciąż rosnących potrzeb automatycznego przetwarzania tekstów i nagrań w językach naturalnych, niezbędnego w wielu dziedzinach przemysłu informatycznego: od wyszukiwania danych aż po tłumaczenia maszynowe, dobrze byłoby wiedzieć, czy nasze możliwości obliczeniowe wystarczą do poradzenia sobie z dowolnymi językami naturalnymi.

Drugim powodem są ogólne pytania teoretyczne o związki pomiędzy tradycyjnym lingwistycznym opisem języków a ich opisem formalnym. Co więcej – wiedząc, jakie własności czysto formalne mają języki naturalne, można próbować orzekać o adekwatności różnych ich opisów gramatycznych.

Trzecią motywacją do badań nad tym zagadnieniem mogą być pytania filozoficzne o sposób działania i możliwości obliczeniowe ludzkiego mózgu. Jeśli bowiem na co dzień używamy konstrukcji, których maszyny nie potrafią przetwarzać w rozsądnym czasie, wówczas można podejrzewać, że mechanizmy odpowiedzialne w mózgu za kompetencję językową wyposażone są w moce obliczeniowe istotnie większe niż moce owych maszyn.

Należy przy tym zwrócić uwagę na jedną rzecz: wszelkiego rodzaju argumenty dowodzące (poprawnie lub nie), że języki naturalne nie są bezkontekstowe, bazują na wskazaniu pewnej konkretnej konstrukcji składniowej w konkretnym języku. Za każdym razem wnioski autorów (sformułowane jawnie lub nie) polegają na tym, że skoro znaleźliśmy jeden język naturalny, którego w całości nie można opisać za pomocą gramatyki bezkontekstowej, to w ogóle dowolny język naturalny nie jest bezkontekstowy. Jest to rozumowanie zakładające uniwersalizm językowy, charakterystyczny dla koncepcji Chomsky'ego i dla generatywizmu w ogóle (por. Mecner 2005). Należy jednak pamiętać, że dla wielu językoznawców, psychologów czy filozofów jest to założenie co najmniej kontrowersyjne. Mimo to rozważania dotyczące klas języków formalnych zdolnych opisać języki naturalne nie są pozbawione sensu, nawet jeśli nie jest się entuzjastą koncepcji uniwersalistycznych.

3. ARGUMENT NIEPOPRAWNY – ZDANIA PORÓWNAWCZE W ANGIELSKIM

We wczesnej fazie dyskusji nad omawianym problemem pojawiło się kilka naiwnych argumentów, które Gazdar i Pullum (Pullum i Gazdar 1982) nazywają folklorem. Omawianie argumentów niepoprawnych nie jest oczywiście celem tego artykułu, ale jeden spośród mniej banalnych tego typu argumentów warto przytoczyć, szczególnie że pochodzi od samego Chomsky'ego.

Jednym z popularnych przykładów sztucznych języków kontekstowych jest tzw. język typu xx (ang. *xx-language*). Ustalmy niepusty alfabet $\Sigma = \{a, b\}$. Językiem typu xx nazywamy język $L = \{xx : x \in \Sigma^+\}$. Oczywiście ciągi x mogą być oddzielone od siebie dowolnym innym ciągiem znaków – to, co czyni taki język kontekstowym, to wymóg wystąpienia dwóch identycznych ciągów znaków z alfabetu (por. Hopcroft i in. 2005). Języki typu xx oraz języki podobne do nich są często stosowane jako argumenty dowodzące tego, że jakies zjawiska w wybranych językach nie są bezkontekstowe.

Noam Chomsky w pracy (Chomsky 1963) formułuje argument na rzecz tezy o kontekstowości konstrukcji składniowej zdań porównawczych w języku angielskim. Chodzi o zdania typu:

- (1) That one is wider than this one is deep.

Jednocześnie za niegramatyczne Chomsky uznaje zdania takie, jak:

- (2) *That one is wider than this one is wide. –

a za jego gramatyczny równoważnik (2) zdanie:

- (3) That one is wider than this one is.

A zatem niepoprawne są te zdania porównawcze, w których następuje powtórzenie tego samego przymiotnika¹, poprawne zaś te, w których przymiotnik występujący w pierwszym członie zdania jest różny od przymiotnika występującego w drugim członie. Chomsky na tej podstawie argumentuje, że konstrukcje porównawcze w angielszczyźnie nie są bezkontekstowe, ponieważ zdania tego typu tworzą język typu xy – język, w którym dwa obowiązkowe człony muszą być różne i który wydaje się łudząco podobny do języka xx . Ustalmy zatem słownik $\Sigma = \{a, b, a, \beta, \gamma\}$. Językiem typu xy nazywamy następujący język:

$$L' = \{\alpha x \beta \gamma y : x, y \in L \wedge x \neq y\},$$

gdzie L jest dowolnym językiem, którego słowa składają się z liter a i b . Chomsky twierdzi, że skoro język L' jest kontekstowy, to również język an-

¹ Przymiotniki różnią się oczywiście stopniem, ale to można w tym wypadku zaniedbać.

gielski zawierający konstrukcje takie, jak w zdaniu (1), musi być kontekstowy. Nie podaje jednak żadnego argumentu na rzecz kontekstowości języka typu *xy*. Jak można sądzić, zakłada milcząco, że jest on kontekstowy z tych samych powodów, dla których język typu *xx* jest kontekstowy. Jednak Gazdar i Pullum (Pullum, Gazdar 1982) zaprezentowali następującą gramatykę bezkontekstową generującą język typu *xy*:

- (4) a. $S \rightarrow \alpha S' \gamma \mid \alpha S'' \gamma$
 b. $S' \rightarrow CS' C \mid D\beta \mid \beta D$
 c. $S'' \rightarrow AB' \mid BA'$
 d. $A \rightarrow CAC \mid a(D)\beta$
 e. $B \rightarrow CBC \mid b(D)\beta$
 f. $A' \rightarrow a(D)$
 g. $B' \rightarrow b(D)$
 h. $C \rightarrow a \mid b$
 i. $D \rightarrow C(D)$

Tym samym języki typu *xy* są bezkontekstowe i argument Chomsky'ego okazuje się nietrafny.

Jak zwracają uwagę Gazdar i Pullum, dla dowolnego języka możemy skonstruować nieskończenie wiele różnych od siebie gramatyk generujących ten język, zatem wartościowy argument na rzecz kontekstowości jakiegoś języka nie może polegać na sprawdzeniu, że wszystkie generujące go gramatyki nie są bezkontekstowe. Jest to uwaga banalna, jednak wiele cytowanych przez nich autorów zdawało się nie dostrzegać tego prostego faktu. Zatem wartościowy argument na rzecz kontekstowości wybranego języka powinien bazować na formalnych własnościach języków bezkontekstowych. Trzy najczęściej stosowane to:

- lemat o pompowaniu dla języków bezkontekstowych,
- domknięcie na homomorfizm,
- domknięcie na przecinanie z językami regularnymi.

Na ostatnim z nich w istotny sposób opierają się najwartościowsze i najczęściej przywoływane argumenty, czyli argument Shiebera ze szwajcarskiej odmiany niemieckiego, argumenty z niderlandzkiego oraz morfologiczny argument Culy'ego z malijskiego języka bambara.

4. SCHWYZERTÜÜTSCH I ARGUMENT SHIEBERA

Schwyzertüütsch (także: Schwyzerdütsch i Schweitzerdeutsch) to grupa dialektów alemańskich języka niemieckiego na terenie Szwajcarii i Księstwa Liechtenstein. Dialekty te dominują w mowie, podczas gdy językiem literackim jest standardowy język niemiecki (jednakże na terenie kantonów Sankt Gallen i Zurych wydawane są książki w miejscowych dialektach alemańskich).

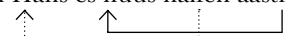
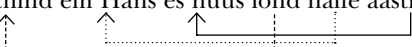
Stuart Shieber w pracy (Shieber 1985) zwrócił uwagę na ciekawe konstrukcje składniowe, występujące w Schwyzertüütsch, a nie występujące w standardowej niemieczyźnie. Zjawisko występuje wyłącznie w zdaniach podrzędnych i dotyczy wymagań składniowych czasowników. Podobnie jak w języku polskim, w Schwyzertüütsch czasowniki, oprócz podmiotu w mianowniku, mogą wymagać również innych fraz w odpowiednich przypadkach gramatycznych – w tym wypadku jedynie w bierniku lub celowniku. Co więcej, niektóre czasowniki wśród swoich wymagań (znów podobnie jak w polszczyźnie) mogą mieć inny czasownik (zwykle w bezokoliczniku). Dzięki temu w zdaniach podrzędnych tworzą się charakterystyczne ciągi czasowników poprzedzone ciągiem ich wymagań walencyjnych w odpowiednich przypadkach. Rozważmy kilka przykładów (ponieważ zawsze rozważane będą jedynie zdania podrzędne, można przyjąć, że każde z nich rozpoczyna się od frazy *Jan säit das...*, czyli *Jan mówi, że...*):

- (5) ...mer em Hans es huus hälfen aastrüiche.
 ...my Janowi-DAT dom-ACC pomogliśmy pomalować.
 ‘...pomogliśmy Janowi pomalować dom’.

W powyższym przykładzie mamy na końcu dwa czasowniki – *hälfen i aastrüiche* – z których każdy wymaga innego przypadku: pierwszy – celownika, drugi zaś – biernika. W Schwyzertüütsch wykładnikami przypadku gramatycznego są takie wyrazy, jak *em* czy *es* z powyższego przykładu.

- (6) ...mer d'chind em Hans es huus lönd.
 ...my dzieciom-ACC Hansowi-DAT dom-ACC pozwoliliśmy
 hälfe aastrüiche.
 pomóc pomalować
 ‘...pozwoliliśmy dzieciom pomóc Janowi pomalować dom’.

W zdaniu (6) znajdują się na końcu trzy czasowniki: pierwszy wymagający biernika, drugi – celownika i trzeci – biernika. Zdania takie można rozbudowywać teoretycznie do dowolnych długości. Przykłady (7) i (8) obrazują graficznie zależności ze zdań (5) i (6).

- (7) ... mer em Hans es huus hälfen aastrüiche.

- (8) ... mer d'chind em Hans es huus lönd hälfe aastrüiche.


Wyjaśnwszy istotę omawianej konstrukcji składniowej w szwajcarskim niemieckim, możemy przejść do zreferowania argumentu Shieberta. Autor

po przedstawieniu danych językowych przyjął, że badany przez niego język ma następujące cztery cechy (jawnie owe cechy wypisując):

– zdania podrzędne w Schwyzertütsch mogą mieć szyk, w którym ciąg wszystkich występujących w nich czasowników jest poprzedzony ciągiem wszystkich występujących w nim fraz nominalnych;

– wśród nich poprawne są również zdania, w których wszystkie frazy nominalne w celowniku poprzedzają frazy nominalne w bierniku oraz wszystkie czasowniki wymagające celownika poprzedzają czasowniki wymagające biernika;

– w zdaniu podrzędnym liczba czasowników wymagających dopełnienia w celowniku musi być równa liczbie fraz nominalnych w celowniku; podobnie dla czasowników wymagających biernika i fraz nominalnych w bierniku;

– w zdaniach podrzędnych takich, jak (5) czy (6), może wystąpić dowolna liczba czasowników.

Niech teraz L będzie językiem posiadającym zaprezentowane powyżej własności (np. Schwyzertütsch). Przypuśćmy, że język ten zawiera zdania takie, jak (5). Weźmy następujący homomorfizm f :

$$\begin{aligned} f(\text{„d'chind”}) &= a \\ f(\text{„em Hans”}) &= b \\ f(\text{„lönd”}) &= c \\ f(\text{„hälfe”}) &= d \\ f(\text{„Jan säit dasmer”}) &= w \\ f(\text{„es huus”}) &= x \\ f(\text{„aastriiche”}) &= y \\ f(s) &= z - \text{w pozostałych wypadkach} \end{aligned}$$

Przecinając język $f(L)$ z językiem regularnym $r = wa^*b^*xc^*d^*y$ otrzymujemy język $f(L) \cap r = wa^m b^n xc^m d^n y$, który nie mieści się w klasie języków bezkontekstowych (jest podręcznikowym przykładem języka kontekstowego, Hopcroft i in. 2005).

Ponieważ języki bezkontekstowe są domknięte na homomorfizm i przecinanie z językami regularnymi (por. Hopcroft i in. 2005), również język L nie jest językiem bezkontekstowym. Zatem język posiadający takie konstrukcje, jak przykład (5), nie jest bezkontekstowy. Na tym kończy się argument Shiebera.

5. POLEMIKA

Shieber w swoim artykule sformułował kilka potencjalnych kontrargumentów, którymi można zaatakować jego oparty na Schwyzertütsch do-

wód, a następnie spróbował je odrzucić. Wydaje się jednak, że zrobił to dość pobieżnie i nie dostrzegł wszystkich stojących za nimi racji.

5.1. WYMAGANIA PRZYPADKA TO NIE SKŁADNIA

Jednym z potencjalnych kontrargumentów, które formułuje sam Shieber, jest twierdzenie, że być może zjawisko wymagania określonego przypadku gramatycznego przez czasownik (na którym to zjawisku opiera się argument ze Schwyzertüütsch) jest nie składniowe, lecz semantyczne. Wówczas oczywiście twierdzenie, że składnię szwajcarskiego niemieckiego można opisać gramatyką bezkontekstową, dałoby się utrzymać. Klóci się to jednak z praktyką badania i opisu badania języków fleksyjnych. Schwyzertüütsch pod względem fleksji bardziej przypomina polszczyznę niż angielszczyznę – zarówno w polskiej, jak i w niemieckiej tradycji lingwistycznej uznaje się wymaganie przypadku za zjawisko składniowe. Nadanie mu interpretacji semantycznej jest oczywiście możliwe, jednak nastęrcza wiele trudności.

W skrajnym wypadku można uznać, że tak naprawdę to, co uznamy za składnię, fleksję, semantykę czy słowotwórstwo, jest kwestią konwencji i wyznawanych poglądów naukowych. Jednak nawet umiarkowani zwolennicy semantycznych interpretacji różnych zjawisk językowych w polszczyźnie nie są skłonni uważać, że wymaganie przypadku jest motywowane semantycznie. Zresztą takie są też odczucia przeciętnego użytkownika języka – zdania, w których wymagania walencyjne czasownika nie są zgodne z przypadkiem gramatycznym frazy w pozycji podmiotu lub dopełnienia, np.:

(9) *Jaś lubi jablek,

nie są uważane za semantycznie dewiacyjne, lecz po prostu za niegramatyczne. Można, bez większych kontrowersji, sądzić, że przeciętny użytkownik szwajcarskiego niemieckiego ma podobne odczucia w stosunku do tego typu zdań w jego języku. Takie zresztą przykłady podaje również Shieber i twierdzi, że wszyscy jego informatorzy jednoznacznie uznali je za niegramatyczne (nie zaś tylko za semantycznie dziwne).

5.2. INNY SZYK

Schwyzertüütsch, podobnie jak polski, jest językiem o dość swobodnym szyku, zatem rozważany układ składników nie jest jedynym układem akceptowalnym. Co więcej, istnieją przesłanki ku temu, by sądzić, że inny szyk tych zdań jest naturalniejszy. Nie jest to jednak argument podważający poprawność samej konstrukcji – faktem jest, że w omawianym języku istnieją poprawne konstrukcje o takiej właśnie budowie i że adekwatna gramatyka takiego języka powinna je opisywać. Argument Shiebera jest niezależny od

tego, czy w Schwyzerdütsch istnieją jakiegokolwiek inne konstrukcje – rozważa on jedynie pewien podzbiór zdań w tym języku i podzbiór ten niewątpliwie nie może być generowany przez gramatyki bezkontekstowe.

Spójrzmy jednak na pragmatyczną stronę całego zjawiska. Argument Shieberta opiera się m.in. na tym, że ciąg czasowników w omawianej konstrukcji będzie uporządkowany pod względem wymagań przypadka – najpierw wszystkie czasowniki wymagające celownika, później zaś – wymagające biernika (lub odwrotnie). Podobnie również uporządkowane muszą być frazy nominalne. Nie wymaga się jednak tego, by fraza nominalna wymagana przez *i*-ty czasownik znajdowała się w ciągu fraz nominalnych na *i*-tym miejscu. Co więcej, Shieber sam podaje przykłady, z których wynika, że frazy nominalne mogą się zamieniać miejscami, nie wymuszając przy tym zamiany miejsc czasowników, które ich wymagają. Jest to zresztą naturalna dla języka z rozbudowaną fleksją swoboda szyku. A skoro tak, to zdania rozważane przez Shieberta są pragmatycznie skrajnie nieskuteczne: mamy bowiem konstrukcje, w których obok siebie występuje kilka fraz celownikowych oraz kilka fraz biernikowych. Liczba możliwych interpretacji składniowych takiego zdania rośnie wykładniczo w stosunku do długości ciągu czasowników w tym zdaniu, albowiem dla każdego znajdującego się w zdaniu czasownika należy rozważyć interpretację z każdą znajdującą się w zdaniu frazą będącą w przypadku gramatycznym, którego wymaga ten czasownik. Zatem w praktyce zdania tego typu nie będą występowały niemal nigdy, bo wielość możliwych interpretacji powoduje, że są one pragmatycznie (choć nie gramatycznie) wadliwe.

5.3. OGRANICZENIE NA DŁUGOŚĆ CIĄGÓW CZASOWNIKOWYCH

Innym kontrargumentem sformułowanym przez samego Shieberta (lecz powtarzonym również później przez innych) jest stwierdzenie, że długość ciągu czasowników w omawianych konstrukcjach jest ograniczona przez pewną stałą liczbę – tym samym można taką konstrukcję realizować za pomocą gramatyk bezkontekstowych. Faktycznie, trudno wyobrazić sobie naturalną sytuację, w której mogłoby się pojawić zdanie z opisywaną konstrukcją zawierające więcej niż pięć czasowników. Jednak rozumując konsekwentnie w ten sposób szybko dojdziemy do wniosku, że dowolne pojawiające się w językach naturalnych konstrukcje, które zwykliśmy opisywać jako zjawiska rekurencyjne i potencjalnie nieskończone, są skończone i ograniczone przez pewne stałe. W skrajnym wypadku uznamy, że istnieje np. z góry ustalone ograniczenie na liczbę zdań prostych, które możemy połączyć spójnikiem współrzędnym. Ale tym samym równie łatwo dojdziemy do wniosku, że języki naturalne, skoro są językami skończonymi, można opisywać za pomocą gramatyk nie tylko bezkontekstowych, ale również regularnych. Przyznać

trzeba, że taka linia obrony bezkontekstowości języków jest trudna do zaakceptowania.

Poza tym oddzielić trzeba dwie rzeczy: czym innym jest adekwatny teoretyczny opis języka, a czym innym – jego implementacja. To przy implementacjach możemy pozwolić sobie na uproszczenia podyktowane ograniczeniami technicznymi, jednak rzetelny naukowo opis teoretyczny powinien być wolny od tego typu wybiegów.

Mimo to należy jednak odnotować kolejne spostrzeżenie pragmatyczne: możliwość wydłużania opisywanych konstrukcji w szwajcarskim niemieckim opiera się na czasownikach o specyficznych wymaganiach walencyjnych. Muszą one móc łączyć się z frazą nominalną w określonym przypadku oraz z innym czasownikiem w bezokoliczniku. W polszczyźnie takich czasowników jest niewiele; można podejrzewać, że w Schwyzertüütsch również stanowią one jedynie znikomy procent wszystkich czasowników. Zatem możliwość tworzenia dłuższych tego typu konstrukcji jest znowu pragmatycznie (choć nie gramatycznie) ograniczona. Warto jednak zwrócić jeszcze raz uwagę (przypominając o ograniczeniach pragmatycznych wzmiankowanych przy jednym ze wcześniejszych kontrargumentów), z jak skrajnie rzadkim zjawiskiem mamy tu do czynienia.

6. INNE PODOBNE ARGUMENTY

W niniejszej części omówię argumenty oparte na dwóch często przywoływanych w kontekście tej problematyki językach: niderlandzkim oraz bambarijskim. Pierwszy z nich pojawił się w tego typu rozważaniach bardzo wcześnie, lecz niektóre oparte na nim argumenty były odrzucane jako nietrafione. Przedstawię jedną z późniejszych wersji argumentu z niderlandzkiego – wszystkie one jednak opierają się na zjawisku krzyżowych zależności znanych już ze szwajcarskiego niemieckiego. Drugi z wymienionych języków, bambarijski, jest ciekawy, ponieważ argumenty z nim związane dotyczą nie składni, lecz morfologii.

6.1. KRZYŻOWE ZALEŻNOŚCI W NIDERLANDZKIM

Jednym z najwcześniej rozważanych na gruncie tego problemu językiem był język niderlandzki. Kilku autorów podawało przykłady konstrukcji dość podobnych do tych ze Schwyzertüütsch, jednak z różnych powodów były one podważane (m.in. w pracy Pullum i Gazdar 1982). Co prawda, argumenty z niderlandzkiego nie wnoszą nic ponad to, co można było zaobserwować na gruncie szwajcarskiego niemieckiego, jednak dla porządku zreferuję krótko jeden z takich argumentów autorstwa A. Manastera-Ramera (Manaster-Ramer 1987).

W niderlandzkim zjawisko krzyżowych zależności występuje, podobnie jak w Schwyzertütsch, w zdaniach podrzędnych, a także w konstrukcjach pytajnych pewnego typu. Opiszę zatem dla odmiany ten drugi wypadek. Oto przykład takiego zdania pytajnego:

- (10) Of Jan Piet Marie zag kussen?
 Czy Jan Piotra Marię zobaczył całować?
 'Czy jak zobaczył Piotra całującego Marię?'

Jak widać, mamy do czynienia z ciągiem czasowników, poprzedzonym ciągiem fraz nominalnych wypełniających wymagania walencyjne tych czasowników. Dodając do tego zjawiska dodatkowo konstrukcję z koordynacją, otrzymujemy następujący przykład:

- (11) Of Jan Piet Marie hoorde ontmoeten en zag omhelzen?
 Czy Jan Piotr z Marią słyszał spotykać i widział obejmować?
 'Czy Jan słyszał jak Piotr spotkał się z Marią i jak [ją] obejmował?'

Istotną cechą powyższego przykładu jest zależność pomiędzy liczbą fraz nominalnych (tutaj: *Jan, Piet i Marie*) a liczbą czasowników w dwóch ciągach czasownikowych (*hoorde ontmoeten* oraz *zag omhelzen*). Mamy tu do czynienia z przecięciem dwóch konstrukcji: krzyżowych zależności pomiędzy czasownikami i ich wymaganiami walencyjnymi oraz koordynacją. Dzięki temu używamy następującą strukturę zdań pytajnych w niderlandzkim:

$NP^n \quad V^n \quad \& \quad V^n$.

Rozważmy zatem następujący język:

$L = \{\text{Of Jan } N^n \text{ Marie hoorde } V^n \text{ ontmoeten zag } W^n \text{ omhelzen}\}$,

taki, że: $N = \{\text{Joop, Alexander, Jan, Wim, Piet, Marie, Willem, Idots}\}$,
 $V = \{\text{horen, zien, helpen}\}$, $W = \{\text{laten, leren}\}$. Łatwo zauważyć, że homomorfizm przeprowadzający N, V, W z L odpowiednio w a, b, c , pozostałe zaś symbole z L w symbol pusty (ϵ) da nam język $a^n b^n c^n$, o którym wiadomo, że jest kontekstowy (a konkretnie jest to język należący do klasy języków indeksowanych, która jest właściwym podzbiorem zbioru języków kontekstowych; por. Hopcroft i in 2005).

6.2. MORFOLOGIA BAMBARYJSKA

Bambara lub bamana to język nigero-kongijski z grupy mande, używany głównie na terytorium Republiki Mali. Liczba mówiących wynosi około 2,7 miliona, dalsze cztery miliony używają bambara jako języka międzyetnicznego w całym Mali. Na podstawie słowotwórstwa bambaryjskiego powstał

inny znany argument autorstwa C. Culy'ego (Culy 1985), pokazujący, że język, który generuje konstrukcje słotwórcze w tym języku (czyli język nad zbiorem morfemów), nie jest językiem bezkontekstowym. Z punktu widzenia dyskusji na temat bezkontekstowości języków naturalnych jest to jednak argument słabszy, ponieważ w teorii składni generatywnej (na gruncie której cały problem został sformułowany) przyjmuje się, że słownik (a zatem wszystkie możliwe konstrukcje słotwórcze w danym języku) jest z góry zadany, zatem w naturalny sposób leży poza sferą zainteresowań badacza konstrukcji zdaniowych czy konstruktora parserów.

Istnieją w języku bambara dwie konstrukcje słotwórcze, na których połączeniu opiera się argument Culy'ego. Pierwsza z tych konstrukcji polega na reduplikacji rzeczownika, dzięki czemu powstaje konstrukcja nieokreślona. Dwa egzemplarze tego samego czasownika przedzielone są morfemem *o*: *N o: N*. Konstrukcja taka znaczy tyle co 'jakikolwiek *N*'.

(12) wulu o wulu
pies pies
'jakikolwiek pies'

(13) malo o malo
ryż ryż
'jakikolwiek ryż'

(14) *wulu o malo
pies ryż

Powyższe przykłady pokazują, że w konstrukcji tej po obydwu stronach *o* musi pojawić się ten sam rzeczownik – inne zestawienia są niepoprawne².

Drugą z interesujących nas konstrukcji bambaryjskich jest konstrukcja agentywna o schemacie: *N + TV + la*, oznaczająca 'ten, który *TV N*'.

Przykłady:

(15) wulu + nyini + la = wuluninyinina
pies szukać
'ten, który szuka psa', 'szukający psa'

² Ogólnie reduplikacja jest bardzo częstym zjawiskiem w morfologii wielu języków (dotyczy zarówno słotwórstwa, jak i fleksji). W powyższym przykładzie mamy do czynienia z pełną reduplikacją (powtórzony zostaje cały wyraz – w przeciwieństwie do częściowej reduplikacji, w której powtórzony zostaje np. tylko jeden morfem), która raczej nie występuje w najbliższych nam językach indoeuropejskich, ale jest zjawiskiem dość częstym i pełni różne funkcje.

- (16) wulu + filè + la = wulufilèla
pies pilnować
'ten, który pilnuje psa', 'pilnujący psa'
- (17) malo + nyini + la = malonyinina
ryż szukać
'ten, który szuka ryżu', 'szukający ryżu'
- (18) malo + filè + la = malofilèla
ryż pilnować
'ten, który pilnuje ryżu', 'pilnujący ryżu'

Wyrazy prezentowane w przykładach (15) i (17) kończą się na *na*, a nie na *la*, ponieważ w bambaryjskim przy określonych zbitkach głosek następuje wymiana *l* na *n*, ale jest to zjawisko motywowane morfologicznie i dlatego nie jest istotne dla argumentu³.

Konstrukcja agentywna jest rekurencyjna, to znaczy wyrazy, które produkuje, można ponownie podstawić jako jej argument⁴:

- (19) wulunyinina + nyini + la = wulunyininanyinina
szukający psa szukać
'ten, który szuka szukających psa'
- (20) wulufilèla + nyini + la = wulufilèlanyinina
pilnujący psa szukać
'ten, który szuka pilnujących psa'

Rzeczowniki uzyskane w drugiej konstrukcji można wstawić do pierwszej konstrukcji:

³ Można to zjawisko porównać np. do polskiego przyimka *w*, który w określonych kontekstach przyjmuje postać *we* – mówimy (i piszemy) *w poniedziałek*, ale *we wtorek*. Nie zmienia to jednak faktu, że *we* jest tylko wariantem przyimka *w*, a nie odrębnym przyimkiem. Decyzja o wyborze tej formy w procesie tworzenia wypowiedzi zapada bardzo późno i jest spowodowana trudną do wymówienia zbitką spółgłosek w wyrażeniu *w wtorek*. Jednak na poziomie słotwórczym czy składniowym zjawisko to nie ma znaczenia.

⁴ Trochę innym, choć równie ciekawym i dającym pewne porównanie ze słowotwórstwem bambaryjskim, zjawiskiem rekursji w morfologii jest tworzenie w polszczyźnie tzw. przymiotników odrzeczownikowych. Mając przymiotniki, np. *międzynarodowy*, *wojskowy*, możemy od nich utworzyć rzeczowniki *międzynarodowość* i *wojskowość*. Od tak uzyskanych rzeczowników w naturalny sposób możemy utworzyć przymiotniki: *międzynarodowościowy* i *wojskowościowy*, a następnie podstawić je do schematu tworzenia rzeczowników odprzymiotnikowych i uzyskać rzeczowniki: *międzynarodowościowość* i *wojskowościowość*. I tylko zdrowy rozsądek powstrzymuje nas przed tworzeniem rzeczowników *międzynarodowościowościowość*, *wojskowościowościowość* i kolejnych.

- (21) wuluninyinina o wuluninyinina
 szukający psa szukający psa
 'jakikolwiek poszukiwacz psa'
- (22) wulunininanyinina o wulunininanyinna
 ten, który szuka ten, który szuka
 poszukiwaczy psów poszukiwaczy psów
 'jakikolwiek poszukiwacz poszukiwaczy psa'

I tak dalej ...

W ten sposób uzyskujemy strukturę podobną do tej, którą rozważał Shieber w swoim argumencie ze Schwyzertüütsch. Niech B będzie słownikiem (zbiorem wszystkich słów) języka bambara (a zatem: zbiorem ciągów morfemów). Niech R będzie zbiorem zdefiniowanym następująco:

$$r = \{wulu(filëla)^h(nyinina)^i \text{ o } wulu(filëla)^j(nyinina)^k : \\ h, i, j, k \geq 1\}$$

Przecięciem B i r jest następujący zbiór:

$$B' = B \cap r = \{wulu(filëla)^m(nyinina)^n \text{ o } wulu(filëla)^m(nyinina)^n : \\ m, n \geq 1\}$$

B' jest postaci $\{a^m b^n a^m b^n : m, n \geq 1\}$, a zatem jest językiem kontekstowym. Ponieważ języki bezkontekstowe są domknięte na przecinanie z językami regularnymi (a r jest regularny), zatem jeśli B' nie jest bezkontekstowy, to i B nie może być bezkontekstowy.

Jak wspominałem już wcześniej, ten argument uważa się za słabszy, ponieważ dotyczy morfologii, a nie składni. Warto jednak zauważyć, że w konstrukcjach słowotwórczych tego typu szyk jest ściśle ustalony, tymczasem w konstrukcjach składniowych opartych na wymaganiach przypadku gramatycznego zwykle dopuszczalny jest również inny układ elementów w zdaniu, albowiem języki o rozbudowanej fleksji posiadają zwykle swobodniejszy szyk.

Jest jeszcze jeden powód, dla którego warto się przypatrzeć argumentowi Culy'ego. Zwraca on bowiem uwagę na ciekawy problem: do jakiej klasy języków z hierarchii Chomsky'ego należy nie tylko składnia poszczególnych języków naturalnych, ale należą również inne podsystemy języka. Próbując krótko opisać to zagadnienie, należy przede wszystkim zauważyć, że poszczególne podsystemy na pewno mają różną złożoność. Najniższy poziom – poziom fonologiczny – jest bez wątpienia najprostszy. Zresztą od początku istnienia nowoczesnego językoznawstwa fonologia uważana jest za część zagadnień języka wyraźnie odrębną od pozostałych. Zajmuje się ona bowiem jednostkami o skończonej i względnie niedużej liczbie, a zatem problemy obliczeniowe, przed którymi staje, są oczywiście z punktu widzenia rozważań

zaproponowanych przez Culy'ego najprostsze. Następny poziom to poziom morfologiczny – fleksja i słowotwórstwo. Na gruncie polszczyzny automatyczna analiza fleksyjna realizowana jest metodami automatów skończonych (a zatem gramatyk regularnych) – problem ten jest zatem rozwiązywany efektywnymi, praktycznie obliczalnymi metodami. Analiza derywacyjna (słowotwórcza) na gruncie polszczyzny jest na razie słabo rozwinięta, ale nic nie wskazuje na to, by miała się w kwestii złożoności obliczeniowej różnić od analizy fleksyjnej. Kolejnemu podsystemowi języka – składni – poświęcona jest większa część niniejszego artykułu. Bez wątpienia analiza składniowa wymaga co najmniej mocy gramatyk bezkontekstowych. Niektóre z zaprezentowanych argumentów pokazują, że istnieją w językach naturalnych konstrukcje, do których analizy gramatyki bezkontekstowe są mechanizmem zbyt słabym. Wreszcie semantyka – ten podsystem języka wydaje się najbardziej złożony, a jednocześnie złożoność jego analizy jest najtrudniejsza do oszacowania. Zatem wydaje się dość ciekawą (choć jednocześnie nie tak zaskakującą) obserwacją fakt, że kolejne szczeble systemu językowego są coraz bardziej złożone.

7. PODSUMOWANIE

W pracy przedstawiono kilka najpopularniejszych argumentów na poparcie tezy, że języki naturalne nie są bezkontekstowe. Ciekawą obserwacją ogólną jest fakt, że odkąd 50 lat temu postawiono ten problem, przez lata niemal bez poważniejszych argumentów na poparcie tej tezy uważano, że języki naturalne wymagają mocy co najmniej języków kontekstowych. Gdy jednak sprawę potraktowano poważnie i rygorystycznie, okazało się, że odpowiedź nie jest tak oczywista. Co więcej, argumenty, które uchodzą dziś za najbardziej przekonujące, pochodzą z języków raczej mało popularnych – wystarczy wspomnieć, że bodaj najbardziej znany argument Shiebera opiera się na konstrukcjach składniowych języka, który właściwie nie istnieje w subkodzie pisanym i rozwija się jedynie jako język mówiony. Zatem nawet jeśli w wypadku ogólnym do opisu języków naturalnych niezbędna jest moc gramatyk kontekstowych, to dotyczy to tak znikomego marginesu wypowiedzeń językowych, że w praktycznych zastosowaniach problem ten można zaniedbać. Kontekstowość w językach jest zatem zjawiskiem niezwykle rzadkim – jej poszukiwaniem zajmowała się przez wiele lat dość spora grupa lingwistów i filozofów, a liczebność zanotowanych przez nich zjawisk jest dość mizerna. Świadczy to raczej o tym, że jest to zjawisko w języku niepożądane – zapewne ze względu na trudności w przetwarzaniu takich wypowiedzi przez nasze mózgi.

BIBLIOGRAFIA

- Chomsky N. (1956), *Three models for the description of language*, [w:] *IRE Transactions on Information Theory*, s.113–123.
- (1957), *Syntactic Structures*. Mouton & Co. Publishers.
- (1963), *Formal properties of grammars*, [w:] R. D. Luce, R. Bush i E. Galanter, [red.], *Handbook of Mathematical Psychology*, vol. II, New York.
- Cully C. (1985), *The complexity of the vocabulary of Bambara*. „Linguistics and Philosophy”, 8, s. 345–351.
- Hopcroft J., Motwani R., Ullman J. (2005), *Wprowadzenie do teorii automatów, języków i obliczeń*, Warszawa.
- Manaster-Ramer A. (1987), *Dutch as a formal language*, „Linguistics and Philosophy”, 10, s. 221–246.
- Mecner P. (2005), *Elementy gramatyki umysłu*, Kraków.
- Pullum G., Gazdar G. (1982), *Natural languages and context-free languages*, „Linguistics and Philosophy”, 4, s. 471–504.
- Shieber S. (1985), *Evidence against the context-freeness of natural language*, „Linguistics and Philosophy”, 8, s. 333–343.